

Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan

Sandi Fajar Rodiyansyah

Abstraksi — Seiring semakin banyaknya ketersediaan data dari hasil aktifitas manusia yang dibantu teknologi informasi, maka diperlukan suatu sistem untuk dapat menghasilkan informasi dan pengetahuan yang terkandung dalam limpahan data tersebut. Teknik data mining memungkinkan untuk menghasilkan pengetahuan dari tumpukan data tersebut yang selama ini tidak dapat diketahui secara manual. Salah satu data yang setiap hari terus bertambah adalah data transaksi penjualan suatu perusahaan dagang. Dengan melakukan data mining pada data transaksi penjualan tersebut diharapkan dapat menghasilkan pengetahuan mengenai pola pembelian oleh konsumen. Pada akhirnya pola yang ditemukan ini dapat digunakan sebagai penjunjang keputusan manajer dalam mengelola aktifitas perusahaannya. Penelitian ini mencari aturan asosiasi menggunakan algoritma apriori.

Kata Kunci—data mining, data transaksi, apriori.

1. Pendahuluan

Data mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, *machine learning* dan basis data. (Han & Kamber, 2006). Data mining digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya data mining, maka akan didapatkan suatu permata berupa pengetahuan (*knowledge*) di dalam kumpulan data-data yang jumlahnya banyak.

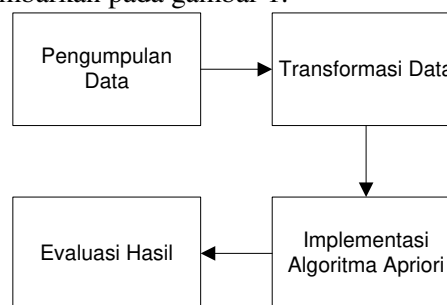
Pada perusahaan perdagangan, terjadi banyak transaksi dan menghasilkan banyak data. Kumpulan data transaksi inilah yang nantinya sebagai informasi yang bermanfaat setelah diolah. Dengan penerapan data mining di dalam pengolahan data tersebut diharapkan dapat membantu proses pencarian informasi yang bermanfaat tersebut.

Dengan demikian, penelitian ini akan mencoba mengimplementasikan data mining dengan algoritma apriori untuk mencari informasi yang berharga berupa pola pembelian barang oleh konsumen dengan memanfaatkan tumpukan data transaksi penjualan.

2. Metode Penelitian

Penelitian ini dilakukan dengan melewati empat tahapan diantaranya : pengumpulan data, transformasi data, implementasi algoritma

apriori dan evaluasi hasil. Tahapan penelitian ini digambarkan pada gambar 1.



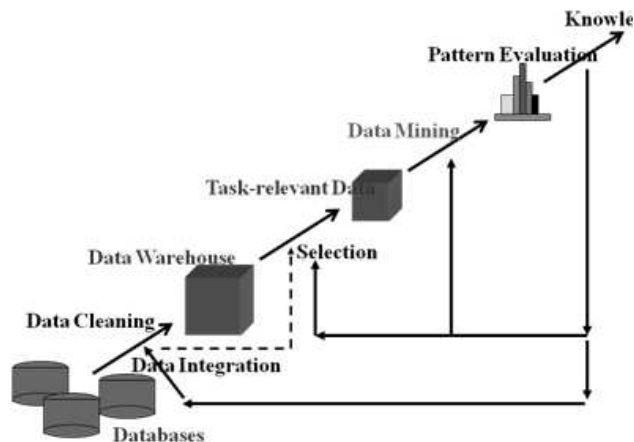
Gambar 1 : tahapan penelitian

Tahap pertama dilakukan pengumpulan data transaksi dari sistem kasir yang telah diimplementasikan sebelumnya. Data yang dikumpulkan adalah data kode transaksi dan barang-barang yang dibeli pada setiap transaksi tersebut. Selanjutnya dilakukan transformasi data. Pada tahapan ini, dilakukan perubahan data transaksi yang telah terkumpul menjadi format data standar untuk proses algoritma apriori. Setelah itu, dilakukan implementasi algoritma apriori dengan menggunakan bahasa pemrograman. Setelah itu, program yang telah dikembangkan digunakan untuk mengolah data yang sudah ada. Pada tahap akhir, dilakukan evaluasi hasil pengolahan data dengan program yang telah dikembangkan.

3. Data Mining

Data mining merupakan suatu metode untuk menemukan pengetahuan dalam suatu tumpukan data yang cukup besar. *Data mining*

adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru dan bermanfaat dan akhirnya dapat ditemukan suatu corak atau pola dalam data tersebut. (Han dan Kamber, 2006). *Data mining* adalah bagian integral dari *knowledge discovery database* (KDD) yang merupakan proses keseluruhan mengubah data mentah menjadi pola-pola data yang menarik yang merupakan informasi yang dibutuhkan oleh pengguna sebagai pengetahuan, seperti yang ditunjukkan pada gambar 2.



Gambar 2 : Proses penemuan pengetahuan dalam database/KDD (Han & Kamber, 2006)

Terdapat empat tugas utama data mining yang terlihat pada gambar 3 yaitu :

1. *Predictive modelling*

Predictive modelling digunakan untuk membangun sebuah model untuk target variabel sebagai fungsi dari *explanatory* variabel. *Explanatory* variabel dalam hal ini merupakan semua atribut yang digunakan untuk melakukan prediksi, sedangkan variabel target merupakan atribut yang akan diprediksi nilainya. *Predictive modelling* dibagi menjadi dua tipe yaitu : *Classification* yang digunakan untuk memprediksi nilai dari target variabel yang diskrit dan regresi yang digunakan untuk memprediksi nilai dari target variabel yang kontinyu.

2. *Association analysis*

Association analysis adalah penemuan association rule yang menunjukkan pola-pola yang sering muncul dalam data. Terdapat nilai *support* dan *confidence* yang dapat menunjukkan seberapa besar suatu rule dapat dipercaya. *Support* adalah ukuran dimana seberapa besar

tingkat dominasi suatu item atau itemset terhadap keseluruhan transaksi. Sedangkan *confidence* adalah ukuran yang menunjukkan hubungan antara dua item secara conditional.

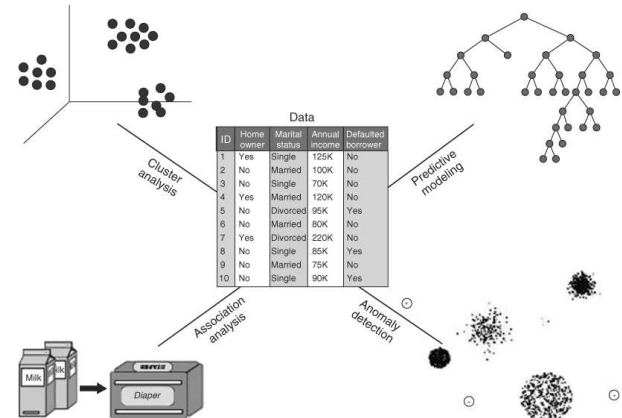
Association analysis digunakan untuk menemukan aturan-aturan asosiasi yang memperlihatkan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

3. *Cluster analysis*

Tidak seperti klasifikasi yang menganalisa kelas data objek yang mengandung label. *Clustering* digunakan untuk menganalisa objek dari data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data *training* karena belum diketahui sebelumnya. *Clustering* merupakan proses mengelompokkan sekumpulan objek yang sangat mirip.

4. *Anomaly detection/outlier mining*

Sebuah database dapat mengandung data objek yang tidak sesuai atau menyimpang dari model data. Data objek ini disebut outlier. Banyak metode data mining yang menghilangkan outlier ini. Padahal, pada beberapa aplikasi seperti *fraud detection*, kejadian yang jarang terjadi tersebut justru lebih menarik untuk dianalisa daripada kejadian yang sering terjadi. Analisa dari outlier data tersebut disebut sebagai outlier mining. Outlier mining juga sering disebut dengan *anomaly detection* yang merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. Anomaly dapat dideteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.



Gambar 3 Empat tugas utama data mining (Tan, dkk, 2006)

4. Algoritma Apriori

Algoritma apriori atau sering disebut juga dengan analisis asosiasi (*association rule mining*) adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item. (Han & Kamber, 2006) Misalkan $I=\{i_1, i_2, i_3, \dots, i_n\}$ merupakan himpunan items dan D merupakan himpunan database transaksi yang setiap transaksi T merupakan himpunan item yang terkandung dalam transaksi T . Setiap transaksi diasosiasikan dengan *identifier* yang disebut TID. Aturan yang memenuhi *minimum support* (*min_sup*) dan *minimum confidence* (*min_conf*) disebut kandidat rule yang kuat. Berdasarkan konvensi, nilai *support* dan nilai *confidence* adalah nilai antara 0% sampai 100% sebanding dengan 0 sampai 1,0 (Linof, 2004).

Himpunan item disebut dengan itemset. Itemset yang mengandung k items disebut k -itemsets. Sebagai contoh himpunan {kopi, gula} merupakan 2-itemset. Kecenderungan kemunculan itemset dalam sejumlah transaksi disebut *frequency*, *support count* atau *count itemset*.

Aturan asosiasi dilakukan dengan melalui dua langkah, yaitu :

1. Temukan semua *frequent itemset*. Hasil dari proses ini adalah sejumlah *itemset* dengan nilai *support* lebih besar atau sama dengan *minimum support* yang diberikan. Nilai *support* dari itemset {A,B} dihitung dengan menggunakan persamaan 1.

$$s = \frac{\text{frequency}\{A,B\}}{\sum T} \quad (1)$$

2. Menentukan *strong rule* dari *frequent itemset*. Hasil dari proses ini adalah semua rule yang memenuhi *minimum support* dan *minimum confidence*. Nilai *confidence* dari suatu kandidat rule $A \rightarrow B$ dihitung dengan menggunakan persamaan 2.

$$c = \frac{\text{frequency}\{A,B\}}{\text{frequency}\{A\}} \quad (1)$$

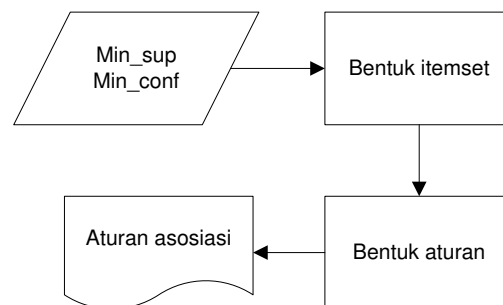
5. Perancangan dan Implementasi

Data yang digunakan dalam penelitian ini adalah sampel data transaksi perusahaan perdagangan dengan struktur data yang tertuang pada tabel 1.

Tabel 1 Struktur data transaksi

| N o | IDT | Item |
|--------|-----|------------------|
| 1 | 001 | Susu, teh, gula |
| 2 | 002 | Teh, gula, roti |
| 3 | 003 | Teh, gula |
| 4 | 004 | Susu, roti |
| 5 | 005 | Susu, gula, roti |
| 6 | 006 | Teh, gula |
| 7 | 007 | Gula, kopi, susu |
| 8 | 008 | Gula, kopi, susu |
| 9 | 009 | Susu, roti, kopi |
| 10 | 010 | Gula, teh, kopi |

Setelah struktur data terdefinisi tahapan selanjutnya adalah perancangan algoritma apriori. Gambar 2 adalah *flow chart* aplikasi analisis asosiasi.



Gambar 4 : flow chart analisis asosiasi

Pada penelitian ini, implementasi dari algoritma dilakukan dengan menggunakan bahasa Java. Gambar 5 merupakan potongan kode program dari sistem ini.

```

1 public class Asosiasi {
2     Koneksi k = new Koneksi();
3     public int jumlahTransaksi() {
4     }
5     public double support(String[][] jmlTrx){
6     }
7     public double confidence(String[][]){
8     }
9 }
  
```

Gambar 5 : Potongan kode program

6. Hasil dan Pembahasan

Gambar 6 merupakan tampilan yang memberikan hasil proses analisis asosiasi untuk data transaksi dengan *minimum support* 30% dan *minimum confidence* 10%. Dengan menggunakan hasil analisis tersebut manajemen perusahaan perdagangan dapat menyusun strategi penjualan yang berkaitan dengan hubungan kemunculan suatu barang secara bersamaan dalam suatu transaksi.

7. Kesimpulan

Pemanfaatan data mining untuk analisis asosiasi data transaksi dapat membantu manajemen perusahaan perdagangan untuk menentukan pola keterkaitan kemunculan barang dalam transaksi penjualan, yang pada akhirnya dapat digunakan untuk menyusun strategi penjualan. Pada penelitian ini telah diimplementasikan analisis asosiasi dengan algoritma apriori untuk menemukan pola pada data transaksi penjualan.

Daftar Pustaka

Han, J., & Kamber, M., 2006, *Data Mining: Concepts and Techniques 2e*, Morgan Kaufmann.

Linof, Gordon S. 2004. *Data Mining Techniques : For Marketing, Sales dan Customer Relationship Management*. Wiley Publishing. Indianapolis.

Tan, P. N., Steinbach, M., & Kumar, V., 2006, *Introduction to Data Mining*, Pearson Education, Boston.